

Single Nucleotide Polymorphisms and Insertion–Deletions for Genetic Markers and Anchoring the Maize Fingerprint Contig Physical Map

I. Vroh Bi,* M. D. McMullen, H. Sanchez-Villeda, S. Schroeder, J. Gardiner, M. Polacco, C. Soderlund, R. Wing, Z. Fang, and E. H. Coe, Jr.

ABSTRACT

Single nucleotide polymorphisms (SNPs) and insertion–deletions (InDels) are becoming important genetic markers for major crop species. In this study we demonstrate their utility for locating fingerprint contigs (FPCs) to the genetic map. To derive SNP and InDel markers, we amplified genomic regions corresponding to 3000 unigenes across 12 maize (*Zea mays* L.) lines, of which 194 unigenes (6.4%) showed size polymorphism InDels between B73 and Mo17 on agarose gels. The analysis of these InDels in 83 diverse inbred lines showed that InDels are often multiallelic markers in maize. Single nucleotide polymorphism discovery conducted on 592 unigenes revealed that 44% of the unigenes contained B73/Mo17 SNPs, while 8% showed no sequence variation among the 12 inbred lines. On average, SNPs and InDels occurred every 73 and 309 bp, respectively. Multiple SNPs within unigenes led to a SNP haplotype genetic diversity of 0.61 among inbreds. The unigenes were previously assigned to maize FPCs by overgo hybridization. From this set of unigenes, 311 (133 SNP and 178 InDel) loci were mapped on the intermated B73 × Mo17 (IBM) high-resolution mapping population. These markers provided unambiguous anchoring of 129 FPCs and orientation for 30 contigs. The FPC anchored map of maize will be useful for map-based cloning, for genome sequencing efforts in maize, and for comparative genomics in grasses. The amplification primers for all mapped InDel and SNP loci, the diversity information for SNPs and InDels, and the corresponding overgoes to anchor bacterial artificial chromosome (BAC) contigs are provided as genetic resources.

MAIZE, one of the most important crops in the world, is a diploid species with an estimated haploid genome size of 2500 Mb, and a high level of sequence complexity due to the abundance of multiple families of repetitive elements (SanMiguel et al., 1996; Myers et al., 2001). Knowledge of the maize genome will allow the improvement of desired traits by refining gene isolation and molecular breeding strategies. Molecular markers provide a means to estimate diversity, to assess genetic relationships, and to map genes underlying important

agronomic traits in crop species. The utility of DNA markers for determining relationships and genetic similarity in maize was reported for DNA markers such as restriction fragment length polymorphism (RFLP) and for simple sequence repeat (SSR) (Taramino and Tingey, 1996; Smith et al., 1997; Senior et al., 1998; Bernardo et al., 2000; Romero-Severson et al., 2001).

Sequence variants of SNPs and/or InDels are the markers of choice for genotyping and mapping because of their abundance and amenability to high-throughput screening. Furthermore, SNPs and InDels can contribute directly to a phenotype (Thornberry et al., 2001) or they can associate with a phenotype as a result of linkage disequilibrium (Daly et al., 2001). Despite the extensive use of SNPs to study human genetic disorders, there have been fewer extensive surveys of SNPs in plants (Tenallon et al., 2001). Previous studies in maize indicated, however, that SNPs and InDels occur at a relatively high frequency in maize genes (Rafalski, 2002; Bhattramakki et al., 2002; Batley et al., 2003a), and in flanking regions of maize microsatellites (Matsuoka et al., 2002; Mogg et al., 2002; Batley et al., 2003b). Single nucleotide polymorphism discovery projects routinely identify InDel polymorphisms that can be used as diagnostic or mapping tools by analyzing size polymorphisms of polymerase chain reaction (PCR) products on agarose gel, as well as SNPs that can be mapped by a number of assay systems (Kwok 2000).

Integrated genetic and physical genome maps are essential for map-based cloning, comparative genomics, and as templates for genome sequencing efforts. A primary goal of the Maize Mapping Project (MMP) is to develop an integrated genetic and physical map of maize to enable the cloning of maize genes controlling phenotypes and crop productivity. Positioning genes to chromosomal location is important for establishing correlation with phenotype, thus facilitating our understanding of biochemical pathways and biological mechanisms controlling agronomically important traits (Davis et al., 1999; Matthews et al., 2001). The MMP has assembled a high-resolution genetic map for the intermated B73 × Mo17 (IBM) population (Lee et al., 2002) consisting of ~1000 RFLP, and ~1000 SSR markers (MaizeGDB, www.maizegdb.org, verified 4 Aug. 2005). Many of the RFLPs and SSRs were generated from EST sequences (Davis et al., 1999; Sharopova et al., 2002), and therefore, directly mark the position of candidate genes for

I. Vroh Bi, Inst. for Genomic Diversity, Cornell Univ., Ithaca, NY 14853; M.D. McMullen, M. Polacco, and E.H. Coe, Jr., Agronomy Dep., Plant Sciences Unit, Univ. of Missouri, Columbia, MO, and USDA-ARS Plant Genetics Research Unit, Columbia, MO 65211; H. Sanchez-Villeda, S. Schroeder, and Z. Fang, Agronomy Dep., Plant Sciences Unit, Univ. of Missouri, Columbia, MO 65211; J. Gardiner, BIO5 Inst., Univ. of Arizona, Tucson, AZ 85721; C. Soderlund, Arizona Genomics Computational Lab., Univ. of Arizona, Tucson, AZ 85721; R. Wing, Arizona Genomics Inst., Univ. of Arizona, Tucson, AZ 85721. Research was performed at the University of Missouri, Columbia, and was funded by NSF Grant DBI# 9872655 to the University of Missouri. Received 6 Dec. 2004. *Corresponding author (biv2@cornell.edu).

Published in Crop Sci. 46:12–21 (2006).
Genomics, Molecular Genetics & Biotechnology
doi:10.2135/cropsci2004.0706
© Crop Science Society of America
677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: BAC, bacterial artificial chromosome; cM, centimorgan; FPC, fingerprint contig; InDel, insertion–deletion polymorphism; MMP, Maize Mapping Project; PCR, polymerase chain reaction; PIC, polymorphic information content; RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

phenotypes and traits. The parents of the IBM population, B73 and Mo17, represent the two major heterotic groups of U.S. maize germplasm, namely Stiff Stalk (B73) and non-Stiff Stalk (Mo17). The IBM population (Lee et al., 2002) includes 302 recombinant inbred lines that underwent four generations of random mating at the F_2 stage. The combination of a large number of lines and the map expansion due to the random mating generations result in a genetic map resource with approximately 17 times the resolving power of the prior maize map standard (Coe et al., 2002).

To develop the resources for physical mapping, three BAC libraries representing approximately 27-fold genome coverage were constructed from the inbred line B73. *Hind*III fingerprinting of all three libraries and contig assembly using FPC software (Soderlund et al., 2000) are being performed at the University of Arizona (Cone et al., 2002; www.genome.arizona.edu/fpc/maize, verified 4 Aug. 2005). The IBM markers have been hybridized to the fingerprinted BACs and the BAC-marker associations have been entered into FPC. This anchors and orders contigs based on the genetic information, hence, providing an integrated genetic and physical map.

Recently, >10 600 maize unigenes were used as the sequence source for overgo probes that were hybridized to high-density BAC filters (Gardiner et al., 2004). These probes identified BACs corresponding to >9300 unigenes. The mapping strategy adopted by the MMP included, *in silico* correspondence, BAC pooling to anchor mapped markers, and the use of SNPs/InDels as a complementary tool to anchor unmapped unigenes has been outlined previously (Cone et al., 2002). Although a number of the FPCs containing overgo hybridizations were located on the genetic map by sequence similarity of the unigene sequences to RFLP or SSR loci on the IBM map, or by the BAC pooling strategy (Yim et al., 2002), the majority of the unigenes (www.agron.missouri.edu/files_dl/MMP/Cornsensus/) are currently without a genetic map position. To speed integration of the maize genetic and physical maps we have undertaken a mapping approach to place unigenes that contain B73/Mo17 polymorphism on the IBM genetic map through SNPs or InDels, thereby anchoring the corresponding contigs.

The goals of the present research are (i) to analyze the utility of InDels identified within the MMP and address how these InDels can serve as general genetic markers for the maize research community; (ii) to conduct physical mapping of unigenes using SNPs and InDels to anchor BAC contigs to the genetic map; and (iii) to disseminate the maize SNP and InDel resources to the scientific community.

MATERIALS AND METHODS

Plant Materials

The genomic DNA from 12 inbred lines was used for SNP and InDel discovery. Among the 12 lines, B73 and Mo17 are the parents of the IBM population (Lee et al., 2002). Four of the lines, Tx303, CO159, T218, and GT119, are the parents of two additional SSR mapping populations (Sharopova et al., 2002). The lines NC7A, Mp708, and Tx501 were chosen to fur-

ther broaden the germplasm examined. The line W22R-scm2 was included because of its extensive use by many maize geneticists. Illinois High Oil (IHO) and Illinois Low Oil (ILO) were added as part of a collaboration project on application to agronomic trait analysis. The SNPs and InDels that were polymorphic between B73 and Mo17 were genotyped in 286 individuals of the IBM population (Lee et al., 2002; Sharopova et al., 2002). Eighty-three maize inbred lines that represent much of the diversity in maize (Remington et al., 2001) were used to assess InDel diversity (Supplemental Table 1). The origin and the composition of the 83 lines are described at <http://www.maizegenetics.net/germplasm/lines.htm> (verified 4 Aug. 2005).

Amplification of Genomic DNA and Sequence Analysis

The unigene set (www.agron.missouri.edu/files_dl/MMP/Cornsensus/, verified 4 Aug. 2005) was generated at DuPont using publicly available maize ESTs to seed their proprietary EST collection (Cone et al., 2002; Gardiner et al., 2004). The 3'-untranslated regions (3'-UTR) of unigenes were targeted to design PCR primers that amplify about 300 to 500 bp, using Primer3 (Whitehead Institute, Cambridge, MA). All PCR reactions were performed using a PTC-225 thermocycler from MJ Research (Watertown, MA). The following touchdown PCR program was used to amplify all the unigenes in the same conditions: 10 cycles of 1 min at 94°C, 1 min at 65°C with -1°C increment per cycle, and 90 sec at 72°C, followed by 35 cycles of 1 min at 94°C, 1 min at 55°C, and 90 sec at 72°C. The PCR amplification of unigenes was confirmed on 2% SFR agarose gels (Amresco, Solon, OH) stained by ethidium bromide.

Large InDels, visible on a 2% SFR agarose gel (Amresco, Solon, OH) between B73 and Mo17, were further analyzed to assess the polymorphic information content (PIC) of each InDel in the 83 inbred lines listed in Supplemental Table 1. The PIC values were calculated using the formula $PIC = 1 - \sum_{i=1}^n f_i^2$, where f_i is the frequency of the i^{th} allele, and n is the number of alleles (Smith et al., 1997).

Loci polymorphic due to InDels detectable on agarose gels between B73 and Mo17 during primary screening were PCR-amplified in 286 individuals of the IBM population, and genotypes were visually determined directly from agarose gels. Primers that did not reveal InDels between B73 and Mo17 and gave consistent, single product amplification in all 12 lines were advanced to the sequencing phase for SNP and short InDel discovery. Nucleotide diversity parameter π (Tajima 1983) and haplotype diversity were analyzed using the DnaSP program (Rozas et al., 2003).

Identification of Sequence Variants and Data Management

A single PCR fragment per unigene was sequenced on both strands. Only SNPs that occurred in at least two of the 12 inbred lines were considered for mapping to minimize sequencing errors and PCR artifacts. Sequence variants were highlighted as described in Fig. 1, and SNP interrogation primers were typically designed with a 50–61°C annealing temperature using the Oligo Analyzer (v. 2.0, IDT, San Jose, CA). Oligonucleotide primers were synthesized by MWG Biotech, Inc. (High Point, NC).

Bioinformatics routines developed to process data in the SNP project required integration of existing programs with custom PERL and Visual Basic scripts for sequence alignment and quality evaluation, automated detection of sequence variants and their position in the sequence, and generation of mapping files (Fig. 1). The SNP discovery and mapping pipe-

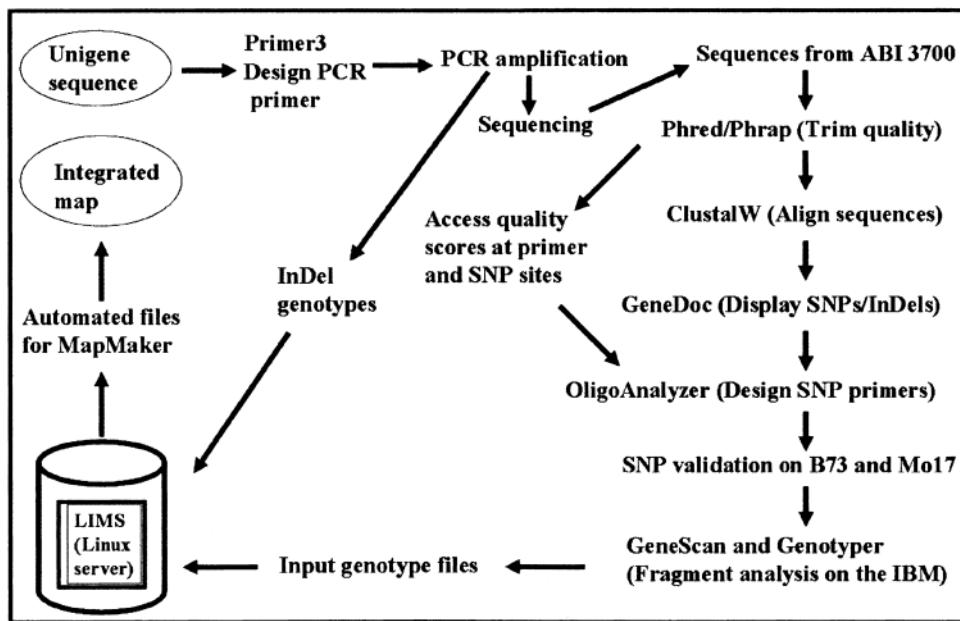


Fig. 1. Pipeline for polymerase chain reaction amplification of unigenes, sequencing, and data processing for genotyping and mapping single nucleotide polymorphisms and insertion-deletion loci.

line was integrated into the Laboratory Information Management System (LIMS) of the MMP (Sanchez-Villeda et al., 2003).

Multiplex Single Nucleotide Polymorphism Assays

The approach for determination of SNP genotypes has seven steps: (i) separate PCR amplification of each unigene, (ii) pooling of up to six different unigene PCR products for simultaneous detection of sequence variants, (iii) exonuclease and phosphatase digestion, (iv) multiplex SNP primer extension reaction, (v) second phosphatase digestion, (vi) resolving labeled interrogation primers on an ABI PRISM 3700 Sequencer, and (vii) analysis of data with GeneScan and Genotyper software (Applied Biosystems, Forster City, CA). Sequence variants detected by sequence alignment were first validated in the parental lines B73 and Mo17 before determining genotypes in the IBM population following the manufacturer's protocol for the SNaPshot Kit (Applied Biosystems, Forster City, CA). The multiplex SNP extension consisted of a 0.5-μL aliquot of the enzyme-treated mix per individual SNP and the extension products were resolved on an ABI PRISM 3700 Sequencer (Applied Biosystems, Forster City, CA).

Linkage Mapping

Genotypes for loci exhibiting SNPs and/or InDels between B73 and Mo17 were determined for 286 individuals of the IBM population. Linkage mapping was conducted using Mapmaker/EXP version 3.0b (Lander et al., 1987) on a UNIX platform. The SNP and the InDel loci were integrated into a framework IBM map consisting of 247 RFLP and SSR loci. Markers were assigned to chromosomes with a minimum LOD score of 15 using the *assign* command. The SNP and InDel loci were ordered on the chromosomes with the *build* (LOD 3) and *place* (LOD 2) commands. Map distances were calculated using the Haldane mapping function.

RESULTS

Optimization of Single Nucleotide Polymorphism Reactions

At the beginning of the genotyping process, it was necessary to optimize reaction conditions and determine

the number of SNPs per multiplex reaction. The DNA of 286 individuals of the mapping population was arrayed into three 96-well plates. After PCR amplification, plates of different unigenes of a given multiplex group were bulked into final plates for enzyme treatment before use in SNP reactions. In a six-plex reaction, this results in avoiding enzyme treatment for 15 plates per multiplex reaction. For the number of SNPs per multiplex reactions, we tested three- to eight-plex. Under our conditions, we found that the maximum number of amplicons per multiplex reaction to achieve a good resolution of SNP peaks on an ABI 3700 capillary sequencer was six. In this case, the maximum volume of multiplex reaction analyzed on the ABI 3700 was 3 μL to keep the background noise at the minimum level. To assess the accuracy of the genotyping and mapping process, the genotype determination process was repeated at three different SNPs, in two different multiplex reactions for three unigenes; AY110160 on chromosome 1, AY110063 on chromosome 5, and AY109644 on chromosome 7. For each unigene, both SNPs mapped to the same genetic location.

Maize SNPs and InDels as Genetic Markers

Genomic regions corresponding to 3000 unigenes were screened by PCR amplification across 12 maize lines. Of these unigenes, 194 (6.4%) showed size polymorphism on agarose gels between B73 and Mo17 due to InDels, herein called large InDels, as opposed to short-length InDels that were present in sequences but did not result in visually obvious size polymorphism on 2% SFR (Amresco, Solon, OH) agarose gels. Accounting for large and short InDels resulted to an average of one InDel every 309 bp. Since the first objective of this research was to anchor the maize FPCs, unigenes showing large InDels between B73 and Mo17 were not sequenced.

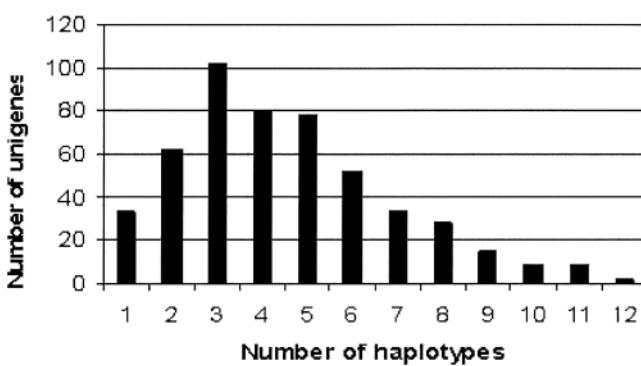


Fig. 2. Distribution of single nucleotide polymorphism haplotype class among 470 maize unigenes analyzed for sequence diversity.

Rather, they were genetically mapped by agarose gel screening of 286 individuals of the IBM population.

We conducted SNP discovery on the first set of 592 unigenes sequenced in the project. The analysis showed that 260 unigenes (44%) contained B73/Mo17 SNPs, while 8% of the unigenes did not show any polymorphism among the 12 inbred lines. On average, 5.5 sequence variants occurred per unigene, with SNPs and InDels occurring every 73 bp and 309 bp, respectively. A previous study in maize found a frequency of one SNP every 70 bp, and one InDel every 160 bp (Rafalski et al., 2001). To assess the utility of SNPs as genetic markers, more detailed analysis was conducted on 470 unigenes having aligned sequence longer than 200 nucleotides across the 12 lines. The genetic diversity (π , the average pairwise difference) assessed in all the 470 unigene alignments varied from 0 to 0.016, with an average of 0.009. Individual SNP are almost always biallelic, leading to a theoretical maximum genetic diversity level of a single SNP marker of 0.5. However, when combinations of SNPs present in the 470 unigenes were used to derive allele haplotypes the genetic diversity range from 0 to 1, with an average of 0.61. This range of SNP haplotype diversity compares favorably with the genetic diversity of RFLP and SSR markers, and demonstrates that SNPs can be highly informative across maize germplasm when analyzed and used at the haplotype level. Unigenes with three haplotypes were the most frequent, with every class from 1 to 12 represented (Fig. 2).

To assess the discriminatory power of InDels as molecular markers, 173 of the 194 large InDels identified between B73 and Mo17 were PCR-amplified in 83 diverse inbred lines (Supplemental Table 1). Figure 3 shows an example of PCR amplification in unigene AY104252

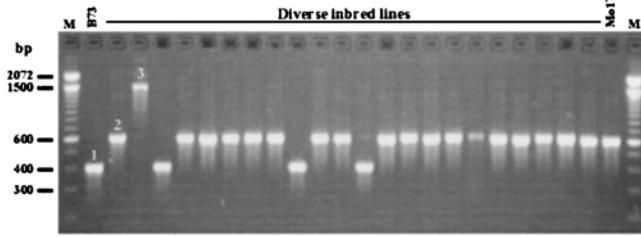


Fig. 3. Profile of a B73/Mo17 insertion-deletion marker in 2% SFR (Amresco, Solon, OH) agarose gel for unigene AY104252 showing three alleles in diverse inbreds of maize.

Table 1. Distribution of allele class for 173 InDel markers tested in 83 diverse inbreds of maize.

Allele class	No. of InDels	Frequency %
2	47	27.1
3	69	39.8
4	42	24.2
5	11	6.3
6	4	2.3

in a set of these inbred lines. The 173 markers detected a total of 539 alleles among the 83 inbred lines. The number of alleles per unigene varied from 2 to 6 with the three-allele class (39.8%) as the largest (Table 1). These results showed that InDels are often multiallelic markers in maize. The PIC values for these InDel markers for the 173 unigenes ranged from 0.04 to 0.76, with an average value of 0.47 (Fig. 4). For assessment of SSRs in the same germplasm, one SSR was chosen per chromosome and tested against the 83 inbred lines. The number of SSR alleles varied from 2 to 5 and PIC values ranged from 0.22 to 0.73, with a mean of 0.54, only slightly higher than that of the large InDel markers. Information on allele number, PIC values, overgo identification, and sequence of PCR primers are given in Supplemental Table 2 for each InDel marker.

Genotyping of SNPs and InDels in the IBM Population

Using one SNP interrogation primer per unigene, we tested 260 unigenes on B73 and Mo17, the parental lines of our IBM mapping population, to validate the SNPs. The B73 and Mo17 genotypes for 246 (95%) SNPs were in concordance with the SNPs predicted by the sequencing data. Among the validated SNPs, 133 were used to determine genotypes in 286 individuals of the IBM population by multiplex primer extension assays. The genotypes of 178 unigenes displaying large InDels between B73 and Mo17 were scored also on the IBM population by direct size polymorphism on 2% SFR (Amresco, Solon, OH) agarose gels. Therefore, a total of 311 unigene loci (represented by 178 InDel and 133 SNP markers) were derived for placement on the IBM genetic map via the LIMS (Sanchez-Villeda et al., 2003) by automated data processing and mapping.

Anchoring FPCs on the IBM Genetic Map

The 311 maize unigenes genotyped in this study were mapped relative to a framework of 247 genetic markers

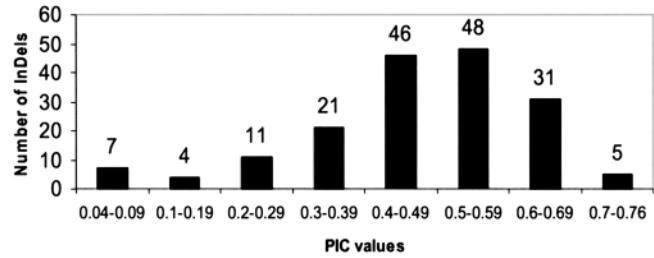


Fig. 4. Distribution of polymorphic information content values for 173 insertion-deletion markers tested in 83 diverse inbreds of maize.

evenly distributed across the 10 maize chromosomes, thus totaling 558 markers (Fig. 5). The genetic map spans 5630 centimorgans (cM), with the number of mapped unigenes ranging from 21 on chromosome 9 to 54 on chromosome 1.

The mapped unigenes were evaluated to determine which might serve to anchor contigs to the genetic map. The 311 unigenes mapped in this study provided unambiguous anchoring of 129 contigs by association with overgo probes used for BAC hybridization (Fig. 5). The unigenes that did not provide unambiguous contig assignment resulted from overgo hybridizations that either failed to identify a contig, or identified multiple contigs (underlined markers in Fig. 5). The multiple contigs identified by a single overgo may represent duplicate gene regions, or contigs that should be merged during manual editing. In 33 cases, two or more adjacent unigenes identified a single contig (Fig. 5), allowing 30 contigs to be oriented relative to the genetic map. The remaining three contigs correspond to off-frame markers, and therefore await the precise placement of these unigenes for orientation. These results allowed a physical-to-genetic ratio to be calculated for 33 regions distributed throughout the 10 chromosomes, from one contig on chromosome 5 and chromosome 9, to seven contigs on chromosome 4. The genetic-to-physical distance varied from 23.1 to 1143 kb cM⁻¹, where the physical distance is measured in the FPC metric CB (consensus band) units, which is approximately 4096 bases per CB. Thus, genetic-to-physical distance showed a variation of almost two orders of magnitude. It is worth noting that the genetic map of IBM is expanded nearly fourfold relative to a standard genetic map (Coe et al., 2002; Sharopova et al., 2002); therefore, these size estimates would be expected to be about four times greater on a standard genetic map.

The amplification primers for all mapped InDel and SNP loci and the interrogation primers for SNP loci are provided in Supplemental Table 3.

DISCUSSION

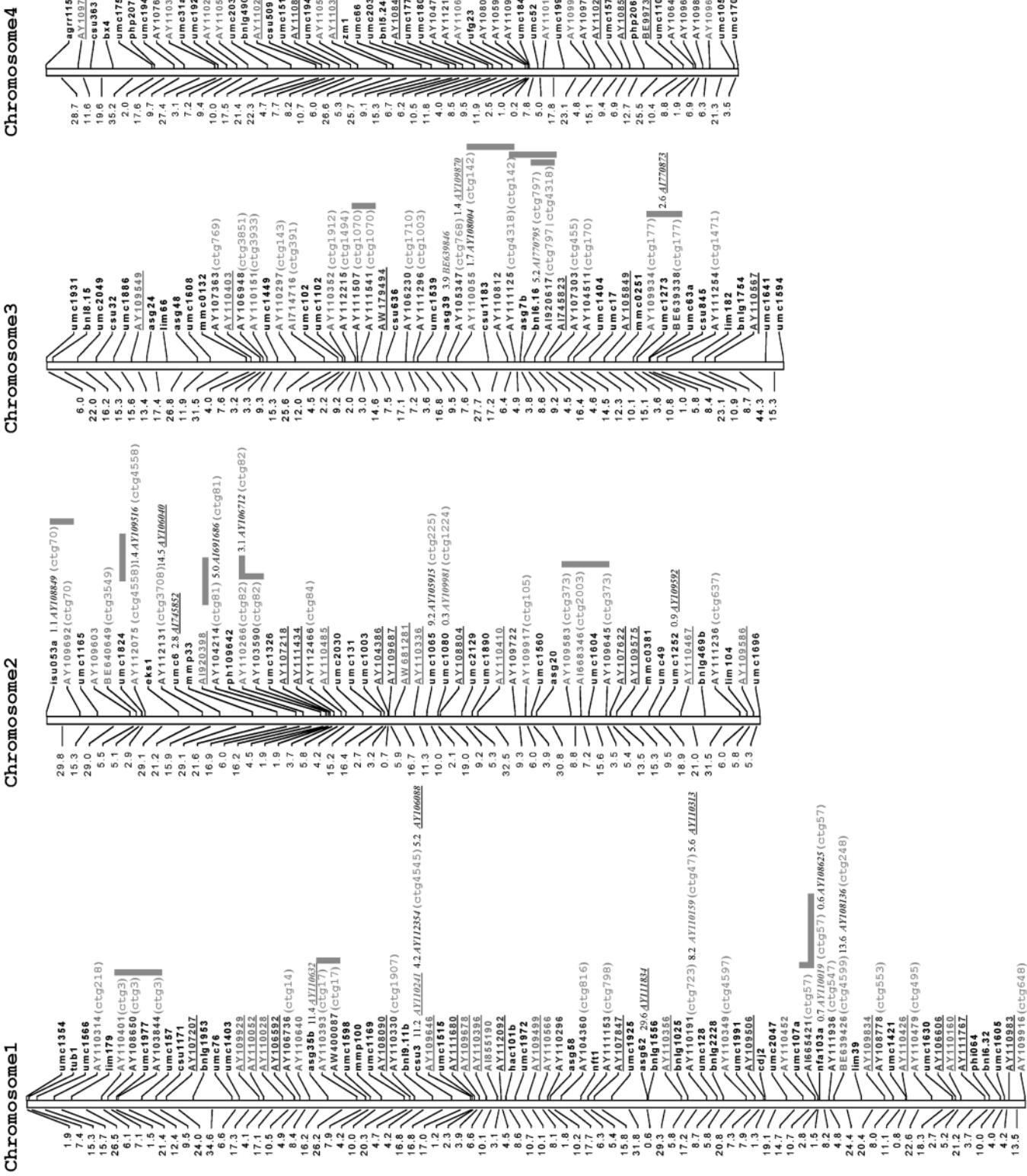
In this study SNP discovery was performed on 592 genes across 12 maize inbred lines. The results confirm those of earlier reports (Tenaillon et al., 2001; Rafalski et al., 2001; Batley et al., 2003a) that both SNP and InDel frequency in maize is very high (one SNP every 73 bp and one InDel every 309 bp). This level of polymorphism is greater than other crop species examined to date such as rice (*Oryza sativa* L.) (one SNP per 268 bp; Shen et al., 2004) and soybean [*Glycine max* (L.) Merr.] (one SNP per 328 bp; Zhu et al., 2003). Although only a single \approx 300 sequence alignment for each gene was obtained, 92% exhibited sequence polymorphism, indicating that an SNP strategy will permit genetic mapping of essentially any maize gene of interest. Although the maize genome contains many duplicate genes and gene families, these did not present a major problem in genotyping assays because the same primers were used for the amplification of genotyping fragments as used in the initial sequencing. Primer sets that amplified multiple genes were detected by the presence of multiple fragments in screening before sequencing or near identical

paralogs were identified by the presence of heterozygous bases within inbred lines. Primer sets yielding either of these problems are excluded from use in SNP genotyping assays.

The integration of genetic and physical maps, which is done by locating the FPCs on the genetic map, is a complex task in eukaryotic genomes such as maize. Through the use of SNPs and InDels as complementary tools to BAC pooling and *in silico* anchoring in the MMP, we have placed 311 unigenes of maize onto the IBM genetic map, among which 293 unigenes (94%) are assigned by overgo hybridization to one or more contigs. Of these unigenes, 167 (54%) correspond with overgo probes that hybridized to fingerprinted BACs, providing unambiguous locations for the corresponding FPCs on the genetic map of maize. Contigs associated with the remaining 126 unigenes were directed toward the manual editing pipeline for anchoring. Only 18 unigenes (6%) failed to identify contigs, probably because of the failure of overgo hybridization. These results clearly demonstrate the power of SNP and InDel analysis to provide genetic markers for maize genetic and physical map integration. The study further shows that in addition to the traditional BAC-end sequencing, SNPs and InDels can be used to orient contigs or to confirm contig orientation on a transcript map.

Thirty-three contigs, anchored by two or more unigenes each, provided the relationship between physical and genetic distance at the anchored points for all 10 maize chromosomes. The present results show that the genetic to physical distance in maize varies more than an order of magnitude. That 311 markers identified 33 pairs of linked markers within the \approx 1600 contigs as of June 2004 suggests that much of the recombination in maize may occur in small physical regions. This agrees with results from studies on recombination rates within the *bronze1* (Dooner, 1986), *anthocyaninless1* (Brown and Sundaresan, 1991), and *waxy1* (Okagaki and Weil, 1997) loci. Because the IBM population underwent successive rounds of intermating, differences in recombination rates between regions should be magnified as each region accumulated recombination events in each generation. Thus, a more nonuniform distribution of recombination is expected for the IBM map than for standard, one-generation maps (Sharopova et al., 2002). The anchoring and ordering of more contigs will give us a comprehensive comparison between the genetic and the physical map distances in maize.

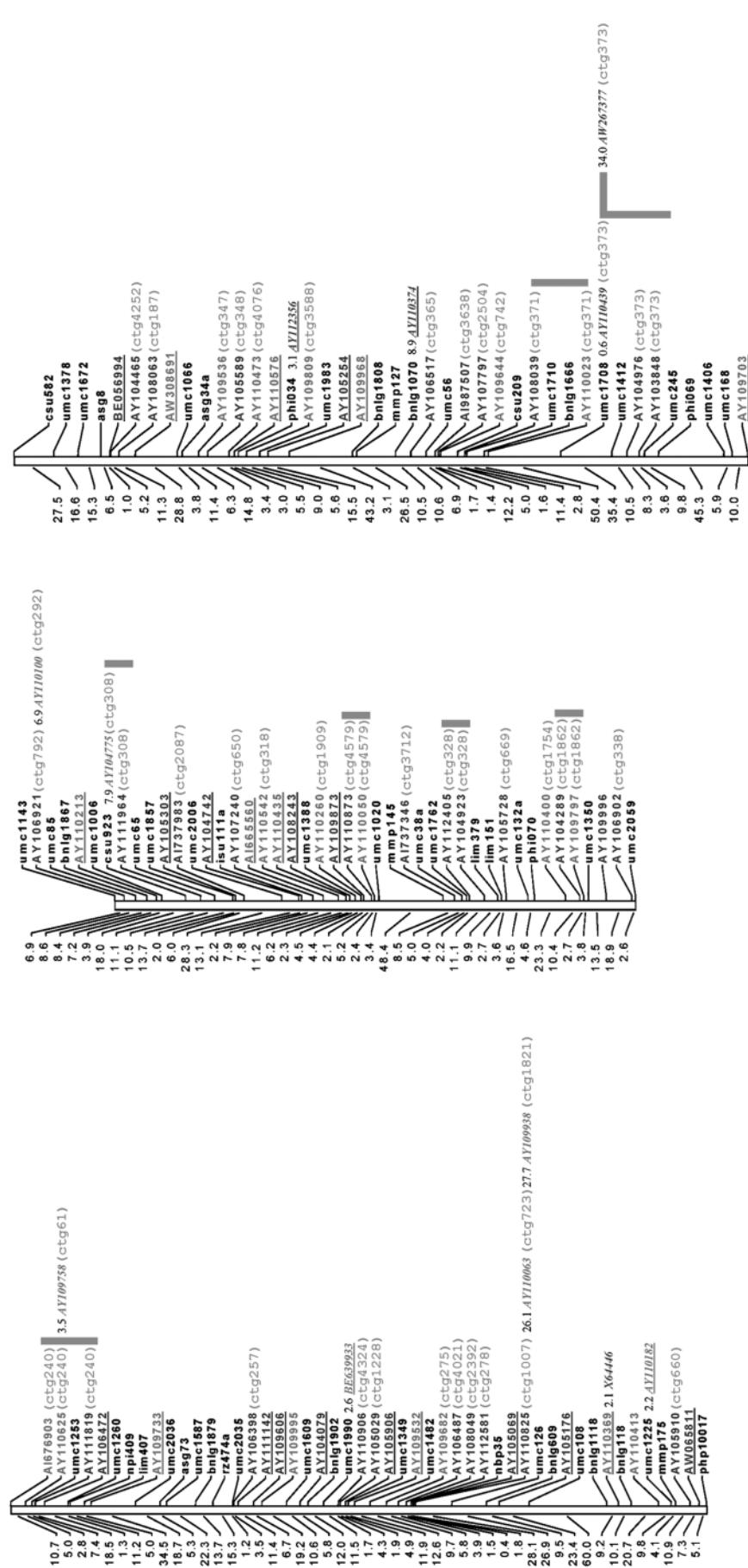
The unambiguous anchoring of FPCs to the IBM genetic map is the main focus of our SNP project. We continue to sequence and map unigenes through SNPs and InDels with the objective of mapping 1000 unigenes to the IBM genetic map. The maize FPC map has now advanced to the state that we can focus our sequencing and SNP mapping efforts solely on unigenes that identify single contigs. This will allow all mapped loci to serve as contig anchor points. These data are being integrated into an iMap for public display www.maizemap.org, verified 4 Aug. 2005). In addition to the detailed information on the SNPs and InDels provided in Supplemental



Chromosome 7

Chromosome 6

Chromosome 5



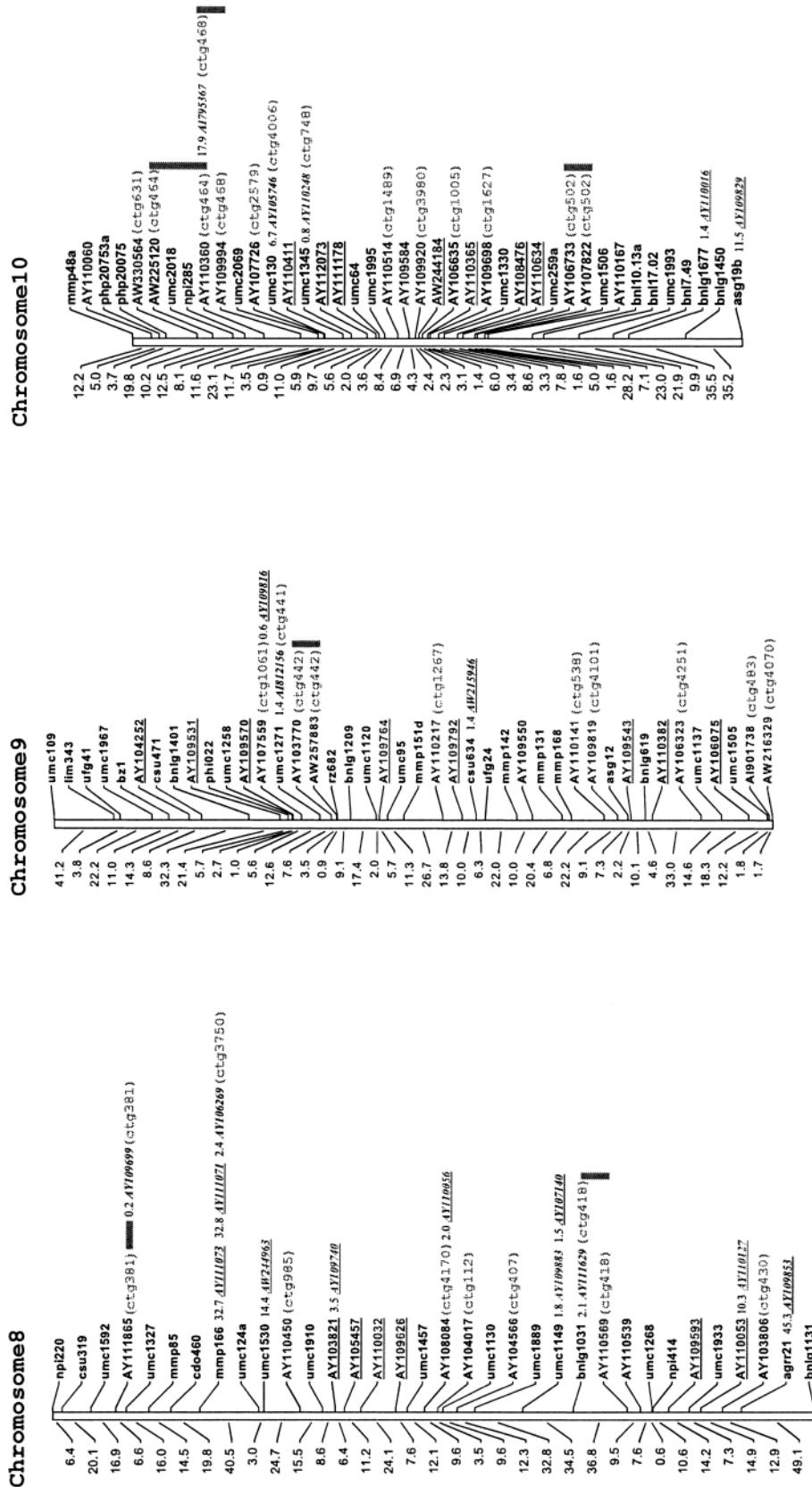


Fig. 5. Integrated genetic and physical map of the intermated B73 × Mo17 (IBM) population based on single nucleotide polymorphisms (SNPs) and insertion-deletion (InDel) markers with the fingerprint contigs of June 2004. The types of markers (SNP or InDel) used to map the unigenes are given in Supplemental Table 3. Bars are bacterial artificial chromosome (BAC) contigs detected by two or more adjacent markers. The BAC contigs associated to unigenes are noted (ctg#). Unigenes that detected several BAC contigs are underlined. Off-frame markers are in italic, and markers that are not listed in Supplemental Table 3 are the framework markers used to assign SNP and InDel markers to maize chromosomes.

- Soderlund, C., S. Humphray, A. Dunham, and L. French. 2000. Contigs built with fingerprints, markers and FPC V4.7. *Genome Res.* 10: 1772–1787.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.
- Taramino, G., and S.V. Tingey. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* 39:277–287.
- Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley, and B.S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl. Acad. Sci. USA* 98:9161–9166.
- Thornberry, J., M.M. Goodman, J.F. Doebley, S. Kresovich, D. Niel-
- sen, and E. Buckler, IV. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* 28:286–289.
- Yim, Y.-S., G.L. Davis, N.A. Duru, T.A. Musket, E.W. Linton, J.W. Messing, M.D. McMullen, C.A. Soderlund, M.L. Polacco, J.M. Gardiner, and E.H. Coe. 2002. Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. *Plant Physiol.* 130:1686–1696.
- Zhu, jY.L., Q.J. Song, D.L. Hyten, C.P. Van Tassel, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young, and P.B. Cre-gan. 2003. Single nucleotide polymorphism in soybean. *Genetics* 163:1123–1134.

Supplemental Table 1. Origin of the 83 diverse inbreds of maize assayed for 173 insertion-deletion polymorphism markers.

Inbred	Origin based subpopulation
B73	U.S. and Northern Stiff
38-11	U.S. and Northern non-Stiff
A272	semitropical inbreds
A441-5	semitropical inbreds
A554	U.S. and Northern non-Stiff
A6	semitropical inbreds
A619	U.S. and Northern non-Stiff
A632	U.S. and Northern Stiff
B103	semitropical inbreds
B104	U.S. and Northern Stiff
B14A	U.S. and Northern Stiff
B37	U.S. and Northern Stiff
B68	U.S. and Northern Stiff
B84	U.S. and Northern Stiff
B97	U.S. and Northern non-Stiff
CI187	U.S. and Northern non-Stiff
CM105	U.S. and Northern Stiff
CM174	U.S. and Northern Stiff
CML247	semitropical inbreds
CML333	semitropical inbreds
CML91	semitropical inbreds
CMV3	U.S. and Northern non-Stiff
D940Y	semitropical inbreds
EP1	U.S. and Northern non-Stiff
F2834T	semitropical inbreds
F44	U.S. and Northern non-Stiff
F7	U.S. and Northern non-Stiff
GT112	U.S. and Northern non-Stiff
H95	U.S. and Northern non-Stiff
H99	U.S. and Northern non-Stiff
HP301	U.S. and Northern non-Stiff
I137TN	semitropical inbreds
I29	U.S. and Northern non-Stiff
IA2132	U.S. and Northern non-Stiff
IDS28	U.S. and Northern non-Stiff
IL101	U.S. and Northern non-Stiff
IL14H	U.S. and Northern non-Stiff
IL677A	U.S. and Northern non-Stiff
K55	U.S. and Northern non-Stiff
KUI11	semitropical inbreds
KUI2007	semitropical inbreds
KUI21	semitropical inbreds
KUI3	semitropical inbreds
KUI43	semitropical inbreds
KUI44	semitropical inbreds
KY21	U.S. and Northern non-Stiff
M162W	semitropical inbreds
M37W	semitropical inbreds
Mo17	U.S. and Northern non-Stiff
MO24W	U.S. and Northern non-Stiff
MS153	U.S. and Northern Stiff
N7A	US and Mixed
N192	U.S. and Northern Stiff
N28HT	U.S. and Northern Stiff
NC250	U.S. and Northern Stiff
NC260	U.S. and Northern non-Stiff
NC296	semitropical inbreds
NC298	semitropical inbreds
NC300	semitropical inbreds
NC304	semitropical inbreds
NC320	U.S. and Northern non-Stiff
NC338	semitropical inbreds
NC348	semitropical inbreds
NC350	semitropical inbreds
NC352	semitropical inbreds
ND246	U.S. and Northern non-Stiff
P39	U.S. and Northern non-Stiff
PA91	U.S. and Northern non-Stiff
SC213R	semitropical inbreds
SC55	U.S. and Northern non-Stiff
SG18	U.S. and Northern non-Stiff
T232	U.S. and Northern non-Stiff
T8	U.S. and Northern non-Stiff
TX601	U.S. and Northern non-Stiff
TZI10	semitropical inbreds
TZI18	semitropical inbreds
TZI8	semitropical inbreds
U267Y	semitropical inbreds
W117HT	U.S. and Northern non-Stiff
W153R	U.S. and Northern non-Stiff
W182B	U.S. and Northern non-Stiff
W64A	U.S. and Northern non-Stiff
WF9	U.S. and Northern non-Stiff

Supplemental Table 3. Continued.

Chrs	Accession	Overgo	Type	Forward primer	Reverse primer	SNP interrogation primer	
AY106075	PC0127444	InDel	tcataatccaggaggaaac	ggctcgagaaaaactatgtatac			
AJ901738	si618009H12	InDel	GCATCCCCATATCCCATGACTAA	CCATGGAAAGCTGGAGATGCTAC	-	AATTAACCTCAAGAGACTTCATATTGAA	
AW16329	si687049E06	SNP	CITGAACCATGCTAACATCATTCC	ATCGGTITTCGAGAAGTACAAAAGA		ACTACCTTCAGAACGAAAGC	
Chr10	AY110060	CL35548_1	SNP	TAGAAAATGGATGACCATCCAAG	TCAGAAGGCCACAGGTATGAAAC		GIGIGITTTAGTTATTTGTGCTA
AW330564	si707029B01	SNP	CGCACATGAGTCGGCTGAAAC	CCACTGGTACTACACAGCATIC	-		
AW225120	si687024F01	InDel	GGGAAATGTTAGTTCTGGCTGA	ACACAAATAAGGGAGTCAAGGA	-		
AY110360	CL2011_1	SNP	TATTTAAGTCGCTTCATGCT	AAGGATACAAAGCTAACAGATGCTG	-	AAATTCCTTGGTACAATTATCTT	
AJ795367	si605011H02	SNP	GCAAGGATCATAATGTTCTCCACA	GAGGGTGATACITGACTGTGAT	-	TCTGTAAAGCTTCAAATGGGAAATA	
AY109994	CL768_1	SNP	GACGAATGTTGTTACAGAAGA	TAIGACTTGCCAAAGGAAATTC	-	GAGGCAAAGGCAGACAGA	
AY107726	PC0062847	InDel	ATGGCAGATGCAAGTGTGAGATGTT	GGATCATATGGATTTCCTCTICA	-		
AY105746	PC0099975	InDel	TTCACACTGATTTGAAGGGTT	ACAAGCCAACATGGCAATGTA	-		
AY110411	CL242_1	SNP	ACTCTACTCTGGCTCAATGTT	ACAACGAGATAATTCGGAGGGAG	-	AAAGAGAGATCGGGAACTGGG	
AY110248	CL4753_-2	SNP	TGICCGITCICCACTTGT	CATIGCTAAACACTGGGICAGTTG	-	AAAAAAAAATAATGGGCTGGGGCA	
AY112073	CL6800_1	InDel	AGATTCCTAGCTTATTCGCC	GACCATTCTGGAGGACTACAGCAG	-		
AY111178	CL5772_1	InDel	CAAGGATCTGACCTTCA	CCCATAGTCUCCATACATAGAA	-		
AY110514	CL948_1	SNP	AAAGAATGCTGAAGAGTTTC	CCCIAACTTCCATAGGTATCTCACAA	-		
AY109584	CL2436_1	SNP	GAGGAGAACAAACTGCTATAGGC	TCTTCTCCATAGGTATCTCACAA	-		
AY109920	CL36143_1	SNP	ATCCAGTGTGATCCCCCTAAATCCTC	CATACTTTGGATGACAGTGTGCG	-		
AY109876	CL1675_1	SNP	GGGCAACGTTCTGTTCTTTACT	TGGAATGGAGACGTGCTTTAAATA	-		
AY106635	PC0086427	InDel	CTAGCTTAGGGCTGCCGTGT	GTGCAAGGGGGAAACATATC	-		
AY110365	CL38215_1	SNP	GGGATGCCACIACAGCTACAACTA	ACTCTATGGCAAGACCATTCAAG	-		
AY109698	CL4149_1	InDel	ACGAGCATCACGAGAACAGG	AAACACACGTAATTACATTTCAA	-		
AY108476	PC0126344	InDel	CGGATTTCACCTTGGTGGAGGA	CCACGTTCTCATCTGGTAAAGG	-		
AY110634	CL3121_1	SNP	TACTACAACGGCTAACGGGTTA	CAGGACCTGCCCTCAAGAAC	-	AAAGCACCTTGCAGCTCATGTT	
AY106733	PC0107634	InDel	ATACCTGCTGGTCTTAGAAGGG	CAGCTCTGTGATGAATAGGAACAAGT	-		
AY107822	PC0087182	InDel	CACITTCGAAATCTTAGGTACAGG	CACAATCCAAATGTGTTACATCCG	-		
AY110167	CL14540_1	InDel	CGACACATCTAGATCAATCCAAGAA	ACTGCTCTGACTCCGACAG	-		
AY110016	CL16508_1	SNP	GGCTCACTGGAAACGAATAAAT	CCAACCCCTGAGCTCAAAAG	-		
AY109829	CL35744_1	SNP	TGACATAAGCTGTCTCAACGTAA	ATGGCACCATATGCCGATG	-		
				CCAATAGTATGCCTACGATAATCATCAA			